Manual of Fisheries Survey Methods II: with periodic updates

Chapter 8: Lake Fish Population Estimates by Mark-and-Recapture Methods

James C. Schneider

TOC

NEXT PAGE

PREVIOUS PAGE

CITATION

Suggested citation:

Schneider, James C. 1998. Lake fish population estimates by mark-and-recapture methods. Chapter 8 *in* Schneider, James C. (ed.) 2000. Manual of fisheries survey methods II: with periodic updates. Michigan Department of Natural Resources, Fisheries Special Report 25, Ann Arbor.

Chapter 8: Lake Fish Population Estimates by Mark-and-Recapture Methods

James C. Schneider

Estimating the actual numbers of fish in a lake is a difficult and time-consuming process for a number of reasons:

1. Populations of fishes in lakes are often extremely large (e.g., bluegills often number in the thousand per acre, most of which are very small); consequently, large numbers of fish must be marked and examined.
2. Any day's sample is likely to include only a very small portion of the total population; therefore, many days of effort may be required to obtain an adequate sample and realize stable ratios of marked to unmarked fish.
3. If sampling takes too long, small fish may grow (recruit) into the size being estimated or marked fish may die at a faster rate than unmarked fish (both cause an overestimate).
4. Fish may avoid sampling, and trap nets and electrofishing are ineffective in deep water (causing an underestimate and a tendency for biologists to study shallow lakes more than deep lakes). Consequently, precision of the estimate depends on random mixing of marked and unmarked fish in areas that can be sampled. Such mixing often occurs in spring or fall.
5. Individual fish may have territories, daily or seasonal movements, or other behavioral patterns which effect vulnerability to sampling.
6. Sampling gear is selective for species and size. Therefore, estimates should be stratified to compensate, then added together as appropriate.
7. There is no certain crosscheck on the accuracy of the population estimate unless known numbers of fish have been stocked or, in the case of reservoirs, the water can be drained and the fish directly counted.

TOC
NEXT PAGE
PREVIOUS PAGE
CITATION

8.1 General procedures

1. Collect a random sample (within gear limitations) of the target species. Nets should be moved every day or every other day to randomly or systematically cover all areas of the lake where the gear is likely to catch fish.
2. Give fish in good condition identifying marks, such as a tag or temporary clip on the tail fin.
3. Tabulate data by species and size group (e.g., inch group).
4. Release fish away from the sampling gear to encourage mixing of marked and unmarked fish.
5. Allow at least 1 day for the marked fish to recover and become mixed.
6. Collect another random sample of fish.
7. Record the ratio of marked to unmarked fish by species and size group.
8. Repeat steps 1-7 until at least 4 recaptures have been made per species-size strata.
9. Adjust as necessary the daily records of marked fish available for fish that die from handling or are removed by anglers. Substantial losses will invalidate the estimate.
10. Calculate for each combination of species and size group (to compensate for gear selectivity), estimates of population abundance (and error) with appropriate formulas.

11. As appropriate, sum size group estimates (and variances) by species to obtain an estimate of the total population (and variance) within the size range actually sampled.

8.2 Variations

In the above general procedure, data is recorded per sampling trip and is summarized on a daily basis. But there are three variations to data collection and analysis:

Multiple-census–Continually mark fish and retain the multiple-sample format throughout the sample period with the goal of utilizing a multiple census formula, such as the Schumacher-Eschmeyer formula.

Bi-census–Plan an initial marking period, a rest period of 1 week or longer to allow fish to recover and mix, then a recapture period, with the goal of utilizing the Chapman modification of the Petersen formula.

Combination–As in the first multiple-census option, continually mark fish but arrange sampling into two periods – a marking period and a recapture period – with the goal of utilizing the Chapman formula. This entails pooling data from several early samples into a combined marking period, allowing a rest period if possible, resuming sampling and changing to a second type of fish mark, then pooling data from several later samples into a combined recapture period. During the recapture period, all unmarked fish are utilized as part of the unmarked catch and only the marks given during the early period count as recaptures (recaptures of the second marks are ignored in computing the ratio of marked to unmarked because they were already counted once as unmarked fish in the second period). Note that all data, derived from both types of marks, also can be used to compute population estimates by a multiple census formula.

TOC
NEXT PAGE
PREVIOUS PAGE
CITATION

The combination approach is recommended because it is flexible. If data from the combination approach are sufficient to calculate a Chapman estimate, then it seems least likely to be biased in case marked fish are not well mixed in 1 day. If data from the study are sparse (as is often the case), they are used most efficiently by a multiple census formula, and the combined approach continually increases the numbers of marked fish available, maximizes number of recaptures, and utilizes all data.

There are some differences in application and interpretation of formulae. The population estimate by the Chapman method applies in the strictest sense to the day marking was completed. Therefore, recaptures obtained even months later can be used to compute an estimate for the last marking date provided marks are not “lost” (as by re-growth of clipped fins or shedding of tags), recruitment into the size group is negligible, and marked and unmarked fish experience similar rates of mortality or loss to emigration. Thus, fish marked in the fall can be recaptured the following spring. Also, fish readily caught and marked during spring spawning runs (such as walleye and northern pike) can be recaptured in early summer when the sexes are more likely to be well mixed. Note that attempts to both mark and recapture spawning fish are quite likely to be biased because males remain on the spawning ground longer than females and fish are freely migrating at that time (i.e., the population being sampled is not “closed”). This bias can be reduced (but not eliminated) by stratifying the data and estimates by sex.

The population estimate by the Schmaucher-Eschmeyer formula is not so closely attributed to one day, but represents the recapture interval, and is most heavily weighed toward the final day. For that reason, try to obtain large samples of fish and reliable ratios on the last day of sampling. One way to accomplish that is to not mark additional fish on the second from last day and pool sample data for the last 2 days.

8.2.1 Chapman variation of Petersen formulas for bi-census

From Ricker (1975); see also Chapter 7:

$$N = \frac{(M+1)(C+1)}{R+1} \quad , \qquad (1)$$

where:

N = population estimate in numbers of fish;

M = number of fish caught, marked and released in first sample;

C = total number of fish caught in second sample (unmarked + recaptures);

R = number of recaptures in second sample (of fish marked and released in first sample).

$$\text{Variance of } N = \frac{(M+1)^2(C+1)(C-R)}{(R+1)^2(R+2)} = \frac{N^2(C-R)}{(C+1)(R+2)} \quad , \qquad (2)$$

$$\text{Standard error } = \sqrt{\text{Variance of N}} \quad ,$$

95% confidence limits of $N = N \pm t$(Standard error) ,

where t is Student's t for $C-1$ degrees of freedom. (See Table 8.1 for t values).

Variance equation (2) should be used whenever variance estimates are to be combined, as for example when summing estimates and variances for two or more size groups to obtain a total population estimate. However, it is not the best estimator of variance for single estimates (Ricker 1975). His recommendation for those is to use either binomial charts or a Poisson distribution (Table 8.1). These provide low and high ranges for R which are then substituted in equation (1) to calculate the lower and upper 95% confidence limits. While 95% confidence limits are often used for research, management can often settle for limits of 68% (±1 standard error).

TOC
NEXT PAGE
PREVIOUS PAGE
CITATION

8.2.2 Schumacher-Eschmeyer formulas for multiple census

From Ricker (1975):

$$N = \frac{\sum_{d=1}^{n} C_d M_d^2}{\sum_{d=1}^{n} R_d M_d} \quad , \qquad (3)$$

where:

N = population estimate in numbers of fish;

$C_d = U_d + R_d$ = total number of fish caught during day d;

U_d = number of unmarked fish caught during day d;

R_d = number of recaptures during day d (of the type of mark under consideration);

M_d = number of marked fish available for recapture at **start** of day d;

d = sample number (usually day), ranging from first (d_1) to last (d_n).

$$s^2 = \frac{\sum_{d=1}^{n}\left(\frac{R_d^2}{C_d}\right) - \left[\frac{\left(\sum_{d=1}^{n} R_d M_d\right)^2}{\sum_{d=1}^{n} C_d M_d^2}\right]}{m-1}, \qquad (4)$$

where:

s^2 = variance of samples;

m = number of days (or samples) in which fish were actually caught.

$$\text{Variance of } N = N^2\left[\frac{Ns^2}{\sum_{d=1}^{n} R_d M_d}\right], \qquad (5)$$

$$\text{Standard error of } N = \sqrt{\text{Variance of } N},$$

$$\text{95\% confidence limits of } N = N \pm t(\text{Standard error}),$$

TOC
NEXT PAGE
PREVIOUS PAGE
CITATION

where Student's t (Table 8.1) is based on $m-1$ degrees of freedom.

Variance equation (5) should be used whenever variance estimates are to be combined, as for example when summing estimates and variances for two or more size groups to obtain a total population estimate with variance. However, as with the Chapman method, it is not the best estimator of variance for single estimates (Ricker 1975). His recommendation is to compute reciprocals of N (i.e., $1/N$) from equation (3) and variances and errors from equation (6) below:

$$\text{Variance of } 1/N = \frac{s^2}{\sum_{d=1}^{n} C_d M_d^2}, \qquad (6)$$

$$\text{Standard error of } 1/N = \sqrt{\text{Variance of } 1/N},$$

$$\text{95\% confidence limits of } 1/N = 1/N \pm t(\text{Standard error}).$$

The reciprocals of those fractional limits are then taken to obtain whole number confidence limits. Note that when reciprocals are taken, the distribution of limits around the point estimate change from symmetrical to asymmetrical. The interval between the point estimate and the lower limit becomes less than the interval between the point estimate and the upper limit.

8.2.3 Alternative methods

The equation for multiple census developed by Schnabel (Ricker 1975) gives estimates very close to those obtained by the Schumacher-Eschmeyer, equation (3).

Depletion methods described in Chapter 7 of Manual of Fisheries Survey Methods II could conceivably be applied to some lake data sets if samples can be arranged into appropriate two-pass or multiple-pass formats. However, the restrictions of this method are more tenuous for lakes than for streams, and mark-recapture methods are usually better in lakes. Restrictions on the depletion method include (a) constant sampling effort; (b) 20% or more of the population is caught per sample (samples may be pooled); and (c) the population is less than approximately 2,000 fish. Constant sampling effort is more feasible in shallow streams, where active electrofishing can thoroughly sample all areas and all fish, than in lakes, where on a given day some fish may choose to avoid passive gear such as trap nets.

8.2.4 Bias

The above formulas provide an estimate of random statistical error but no measure of bias. Errors from bias can be much larger and more serious. Bias is very difficult to determine unless the fish population is known or can be logically bracketed. For example, if a lake was carefully stocked with known numbers of fingerling walleyes, the number of survivors estimated to be present at a later date obviously cannot exceed the number stocked and should decline progressively due to natural and fishing mortality. Likewise, the number of fish in each year class must progressively decline each year due to mortality.

TOC
NEXT PAGE
PREVIOUS PAGE
CITATION

Bias can be introduced by either uncontrollable fish behavior or by failure to use the best procedures. Bias due to fish behavior includes “trap-happy” or “trap-shy” tendencies, territoriality or other distribution tendencies, and any other behavior which can cause non-random samples. Bias can also be introduced by failure to distribute marked fish fairly, sample the whole lake, move nets frequently, correct for loss of marked fish, stratify by species and size to compensate for gear selectivity, or any other procedural flaw which can cause non-random samples. Sometimes, behavior and distribution bias can be compensated for by using one type of gear to collect fish for marking and another for recapture. This works to the extent the gears have different types of bias, but it requires that the target species and size be vulnerable to both types of gear and that a large proportion of the population be handled to obtain tight confidence limits. Usually, random statistical errors are so large they preclude the ability to confirm the presence of bias errors.

At best, our estimates are approximations of numbers of fish present. The most trustworthy statistic is the number of fish actually handled during the procedure; it provides the minimum population size.

Example 8.1–Jewett Lake is a small (12.9 acres), shallow (16 feet), landlocked lake containing only bluegill, yellow perch, and walleye. Population estimates for each species and size were made for many years during a study of population and community dynamics. In the fall and spring of the year, when water temperatures were 55-65°F, large fish were readily collected with regular trap nets (RTN), and medium and small fish were sampled with small-mesh trap nets (STN) and electrofishing (EF). At such cool water temperatures, few fish were harmed by handling. Catches were usually much higher on the first day of netting, suggesting that marked fish may become less active and less vulnerable for a couple of days. Consequently, the lake was sampled the last week in September and the second week in October with a combination plan. If catches were large, a Chapman estimate was calculated; if relatively few fish were marked or recaptured, a Schmacher-Eschmeyer estimate was made. In the first week of sampling, fish were marked by clipping the top lobe of the caudal fin; during the second week of sampling fish were marked by clipping the bottom lobe of the caudal fin. A better procedure statistically would have been to give a unique fin clip for each of the three types of fishing gear; however, six different clips would have been required and the fish would have been unduly stressed. The following table was set up and filled out daily for each species and inch group to aid in tracking progress towards obtaining enough recaptures and for computing population estimates and CPE:

		6-inch bluegill						**7-inch bluegill**		
Date	Gear	*M*	*U*	R_t	R_b	R_{tb}	Notes	*M*	*U*	etc.
9/21	3RTN		55				1 U dead	etc.		
	2STN		13							
	2 hr ES		39							
	Total	0	107							
9/22	3RTN		23	2						
	2STN		8	0						
	2 hr ES		26	1						
	Total	106	57	3						
9/23	3RTN		18	3						
	2STN		11	2						
	1 hr ES		5	0						
	Total	162	34	5			1 R_t DOB			
End 1st week		196								
10/6	3RTN		16	2						
	2STN		2							
	3 hr ES		32	4						
	Total	196	50	6						
10/7	3RTN		20	2	2					
	2STN		4							
	2 hr ES		7	1		1				
	Total	246	31	3	2	1				
Combined 10/6-7			81	9	2	1				
Min. population		277								

TOC
NEXT PAGE
PREVIOUS PAGE
CITATION

Chapman-Petersen Method

$$N = \frac{(196+1)(81+9+1)}{9+1} = 1{,}793$$

$$\text{Variance of } N = \frac{(196+1)^2(81+9+1)(81+9-9)}{(9+1)^2(9+2)} = 285{,}740$$

$$\text{Standard error of } N = \sqrt{285{,}740} = 534.5$$

Symmetrical 95% limits = 1,793 ± 2(534.5) = 1,793 ±1,069 = 724 to 2,862 fish

Asymmetrical limits:

Poisson limits for R=9 are 4.0 and 17.1 (Table 8.1).

Substituting those for R in the N formula above gives 95% limits of 990 to 3,585 fish.

Schumacher-Eschmeyer method

Basic calculations: $\sum RM = 3{,}780 \quad \sum CM^2 = 6{,}088{,}064 \quad \sum R^2/C = 2.4068 \quad m = 5$

$$N = \frac{6{,}088{,}064}{3{,}780} = 1{,}611$$

$$s^2 = \frac{2.4068 - \left[3{,}780^2/6{,}088{,}064\right]}{5-1} = 0.0149618$$

$$\text{Variance of } N = 1{,}611^2\left[\frac{1{,}611(0.0149618)}{3{,}780}\right] = 16{,}549$$

Standard error of N = 128.6

Symmetrical 95% limits of N = ±2.776(129) = ±357 = 1,254 to 1,968

Asymmetrical limits:

$$1/N = 1/1{,}610 = 0.0006207$$

$$\text{Variance of } 1/N = \frac{0.0149618}{6{,}088{,}064} = 2.4576\text{E-09}$$

$$95\% \text{ limits of } 1/N = \pm 2.776\sqrt{2.459E-09} = \pm 0.00013762 = 0.000483 \text{ to } 0.000758$$

Reciprocals: Lower 95% = 1/0.000758 = 1,319; Upper 95% = 1/0.000483 = 2,070

TOC
NEXT PAGE
PREVIOUS PAGE
CITATION

Explanation of Example 8.1 and calculations

On 9/20, nets were set. On 9/21, net lifting, electrofishing, and marking of top tails began. On 9/21, one fish in RTN was in poor condition and was not marked and released; therefore, *M* available for 9/22 was adjusted to 107-1=106. On 9/22, top tail fish were officially available for recapture. On 9/23, one R_t was found dead on the beach and subtracted from *M* available for that day (106+57-1=162). On 9/23, nets were pulled, and by the end of the 1st week total M_t available was 196. A rest period occurred 9/23 to 10/4 to allow mixing and resumption of normal behavior. On 10/5, nets were reset. On 10/6, net lifting, electrofishing, and marking of bottom tails began. On 10/7, it first became possible to collect R_b and R_{tb} clips as well as the original R_t clips. Unmarked fish caught on the 10/7 were not marked because it was anticipated that would be the last day of sampling.

For the Chapman estimate, a spreadsheet was setup for the computation, where *M* = marked fish available after the first week (196); *R* = total recaptures of those fish during the second week (R_t=9); *U* = total unmarked fish in the second week (81), and *C* = *U*+*R* (=90). Note that R_b and R_{tb} are not used because those fish contributed to the ratio the first time they were caught during the second week. Student's *t* value for 90-1 degrees of freedom is essentially 2.0 (Table 8.1).

For the Schumacher-Eschmeyer estimate, a spreadsheet was setup to compute for each strata (combination of species and size) the intermediate statistics of RM, CM^2, and R^2/C for each day and their sums. Then the population estimates and both symmetrical and asymmetrical limits were computed. Note that *M* refers to the number of marked fish available at the **start** of the day's sampling – it does not include fish marked and released that day – and for this estimate includes all three types of fin clips (total of 246 by start of 10/7). Likewise, recaptures of all three fin clips count as recaptures (total of 20 for the entire period). Note that *m*, the number of days catches were made, was 5, and Student's *t* value for 5-1 degrees of freedom is 2.776 (Table 8.1).

For either formula, if recaptures for this strata (6-inch bluegills) had been less than 4, then data from adjacent strata, such as 5-inch bluegills, should have been pooled and a combined estimate calculated. Then, if necessary, the combined estimate could be apportioned between 5- and 6-inch groups according to catches by the least bias gear (probably electrofishing in this study).

For either formula, if estimates from two or more strata are to be combined, the symmetrical variances are to be used. For example, adding an estimate of 800 with a variance of 120,000 to the Jewett example yields for the Chapman method a combined estimate of 2,593 (800+1,793), a combined variance of 405,740 (120,000+285,740), and a combined 95% confidence limit of ±1274 (2 times square root of 405,740); and for the Schumacher-Eschmeyer method a combined estimate of 2,410 (800+1,610), a combined variance of 136,534 (120,000+16,534), and a combined 95% confidence limit of ±739 (2 times square root of 136,534).

Multiple-pass depletion methods were also applied to the Jewett Lake example for comparison (Chapter 7, formula 7). Only unmarked daily catches in trap nets were used. The resulting estimate was 202 bluegills, which was far too low since the population was known to exceed 277 bluegills. The plot of catch rate per sample had considerable scatter around the regression line, indicating the requirement of constant daily catchability was not met. Therefore, the depletion method was not a good choice for the Jewett Lake data set.

Which of the mark-and-recapture results are the best? We know for sure that at least 277 6-inch bluegill were present because that many different fish were handled. The population estimates differ by 11% (1,610 versus 1,792 fish), with the Schumacher-Eschmeyer result being lower (as usual). The Schumacher-Eschmeyer result has tighter confidence limits (asymmetrical: 1,318 to 2,069, a range of 751 fish; symmetrical: 1,252 to 1,968, a range of 716 fish) than the Chapman-Petersen (Poisson: 990 to 3,585, a range of 2,595 fish; formula (2): 724 to 2,862, a range of 2,138 fish) because more recaptures are utilized (20 versus 9). On the other hand, the multiple census method has a greater potential for bias if marked fish did not resume random behavior in 1 day. Some readers may find it disconcerting that the two methods for calculating confidence limits produce such different results. Keep in mind that both point estimates and their error bounds are

TOC
NEXT PAGE
PREVIOUS PAGE
CITATION

but approximations. The choice of which numbers and methods to accept may be influenced also by (a) the need to statistically combine estimates for other strata; (b) the desirability of maintaining consistent methodology across strata and years; and (c) evidence for bias as indicated by unreasonable trends in year class estimates across successive years.

Table 8.1.–Poisson distribution of lower and upper 95% confidence coefficients[a] for number of recaptures (*R*), and Student's t values (∝=0.05) for number of degrees of freedom (df).

Poisson distribution						Student's *t* value	
R	Lower	Upper	*R*	Lower	Upper	df	t_{95}
0	0.0	3.7	26	17.0	38.0	1	12.706
1	0.1	5.6	27	17.8	39.2	2	4.303
2	0.2	7.2	28	18.6	40.4	3	3.182
3	0.6	8.8	29	19.4	41.6	4	2.776
4	1.0	10.2	30	20.2	42.8	5	2.571
5	1.6	11.7	31	21.0	44.0	6	2.447
6	2.2	13.1	32	21.8	45.1	7	2.365
7	2.8	14.4	33	22.7	46.3	8	2.306
8	3.4	15.8	34	23.5	47.5	9	2.262
9	4.0	17.1	35	24.3	48.7	10	2.228
10	4.7	18.4	36	25.1	49.8	11	2.201
11	5.4	19.7	37	26.0	51.0	12	2.179
12	6.2	21.0	38	26.8	52.2	13	2.160
13	6.9	22.3	39	27.7	53.3	14	2.145
14	7.7	23.5	40	28.6	54.5	15	2.131
15	8.4	24.8	41	29.4	55.6	16	2.120
16	9.2	26.0	42	30.3	56.8	17	2.110
17	9.9	27.2	43	31.1	57.9	18	2.101
18	10.7	28.4	44	32.0	59.0	19	2.093
19	11.5	29.6	45	32.8	60.2	20	2.086
20	12.2	30.8	46	33.6	61.3	21	2.080
21	13.0	32.0	47	34.5	62.5	22	2.074
22	13.8	33.2	48	35.3	63.6	23	2.069
23	14.6	34.4	49	36.1	64.8	24	2.064
24	15.4	35.6	50	37.0	65.9	60	2.000
25	16.2	36.8				∞	1.960

[a] Substitute the coefficients for *R* in formula (1). For larger values of *R*, use the following equation (Ricker 1975) for 95% limit coefficients: $R + 1.92 \pm 1.96\sqrt{R + 1.0}$.

8.3 References

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Fisheries Research Board of Canada, Bulletin 191.

Written 3/1947 by W .R. Crowe.
Revised 3/1976 by G. P. Cooper and J. R. Ryckman.
Rewritten 8/1998 by J. C. Schneider

TOC
NEXT PAGE
PREVIOUS PAGE
CITATION

TOC
NEXT PAGE
PREVIOUS PAGE
CITATION